le : 26.09.2017 :

⇒ 1) Population = Ensemble d'éléments (Individus) auquel on s'intéresse à 1 ou à plusieurs caractères.

ex : Un champs de blé de 20 ha → Ensemble des plants.

(Choix des engrais)

(Analyse de la variance)

(permet de comparer plus de 03 moyennes.

Si 2 → Comparaison de 02 moyennes.

+ Analyse en composantes principales (ACP).

⇒ la variance : c'est la moyenne des variantes.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n} , \quad \sigma = \sqrt{\sigma^2}$$

le : 02.10.2017 :

Rappels de statistiques

1- Population : Désigne l'ensemble des individus auquels on s'intéresse à l'1 ou plusieurs de leurs caractères.

ex : Une parcelle de 20 ha de blé

population = ensemble des plants de blé

Individus, Unité, éléments.

2- Echantillon : c'est une fraction de la population sur laquelle porte les observations.

L'objectif est à partir de la connaissance acquise sur l'échantillon

→ faire des déductions correctes sur la population origine.

"(Il faut que l'échantillon soit représentatif)" :

- Echantillonnage aléatoire.
- Effectif suffisamment grand.

(Comment procéder à l'échantillonnage) ???

3- Caractère : (les caractères c'est pour vérifier les hypothèses).

02 types de caractères :

3.1. Caractère Quantitatif = on peut lui attribuer un valeur numérique

ex: longeurs des épis ; Nbre de feuilles / plante ...

① caractère quantitatif continu : Quand il peut prendre théoriquement toutes les valeurs comprises dans un Intervalle (Poids, taille, vitesse...).

② Caractère quantitatif discontinu ou discret : Il prend des valeurs Isolées, valeurs entières (Nbre, ...).

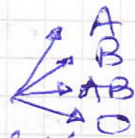
3.2. Caractère Qualitatif :

ex: forme des tubercules, couleur des fleurs, Groupe sanguin, ...

le : 04.10.2017 :

→ un caractère qualitatif présente ≠ modalités.

ex: le groupe sanguin comporte (04) modalité



si le caractère comporte uniquement (2) modalités —dit→ le caractère est dichotomique.

ex: sexe → Masculin / → féminin ex₂ : Marche ou Panne.

→ Cas d'1 caractère quantitatif : le caractère est variable statistique X qui prendra les valeurs $x_1, x_2, x_3, \ldots, x_n$

ex: le Nbre d'enfants / famille = variable statistique X

X	
1	
2	
3	
.	
n	

4- Recensement et Sondage :

- on procédera à un recensement dans une population de taille N, si l'on examine les N individus de cette population selon le caractère retenu.

- On procédera à un sondage dans une population lorsque l'on prélèvera un nombre n d'individus dans cette population ;

cette fraction de la population constitue un échantillon de taille ou d'effectif n.

ex: on fera un sondage sur 500 livres parmi les 50.000 livres d'1 bibliothèque. N' = 50000, n = 500.

5- les Notations :

Paramètre ou caractéristiques	Population	Echantillon
Effectif	N'	n
Moyenne	μ (mu)	$\bar{x}$
Variance	σ^2	s^2
Ecart-type	σ	s
(Population) proportion	π	p
Coefficient de corrélation	ρ	r

6- Rappels de définitions :

- moyenne = $\bar{x} = \frac{\sum x_i}{n}$

$$\bar{x} = \frac{\sum n_i x_i}{n}$$

- Variance = $s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$

$$s^2 = \frac{\sum n_i (x_i - \bar{x})^2}{n}$$

- Ecart-type = $s = \sqrt{s^2}$.

Classes	n_i	x_i	$n_i x_i$
5 - 10	2	7,5	
10 - 15	8	12,5	
20 - 25	10	11,5	
30 - 35	25	⋮	
	$\sum n_i = n$		$\sum n_i x_i$

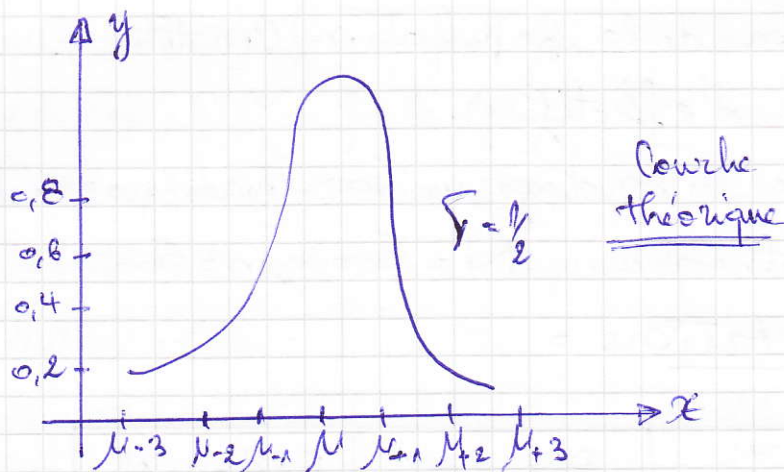
7- Rappels des principales distributions théoriques :

7.1. la loi normale ou la loi de GAUSS :

C'est une loi capitale aussi bien pour le calcul des probabilité qu'en statistiques.

L'expression analytique de la densité :

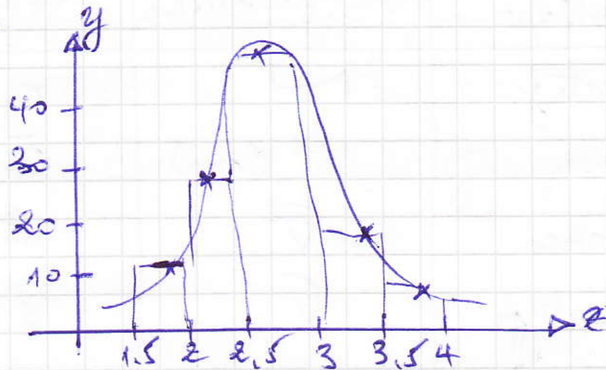
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$



Courbe théorique

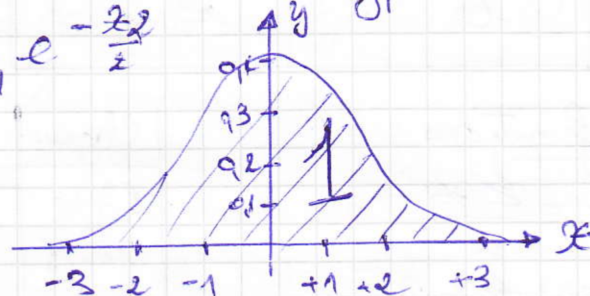
ex : Poids de tubercules.

classes	n_i
1,5 - 2	10
2 - 2,5	18
2,5 - 3	35
3 - 3,5	40
3,5 - 4	12



→ Une loi normale centrée réduite est une loi normale de moyenne $\mu = 0$ et un écart-type $\sigma = 1$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



- Elle est symétrique / Oy (axe des y).
- $\int_{-\infty}^{+\infty} f(x)\, dx = 1/\sqrt{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1$ ou bien : $\int_{-\infty}^{+\infty} f(x) = 1$

elle donne : $Pr(X \leq x_0) = \int_{-\infty}^{x_0} f(x)\, dx$ → (table de la loi normale)

ex : $Pr(X > x_0)$, $Pr(X < -1)$.

le : 09.10.2017 :

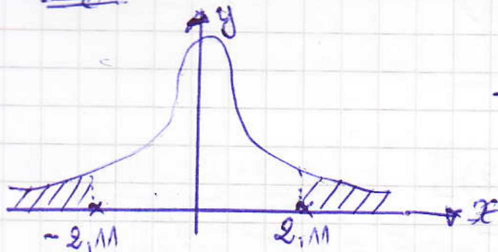
ex 01 = $Pr(X \leq 2{,}11) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{2{,}11} e^{-\frac{x^2}{2}} dx = 0{,}98257$ ⇒ $(2{,}11 = 2{,}1 + 0{,}01)$ table de loi normale



	0,00	0,01
0,0		
1,1		
2		
2,11		0,98257



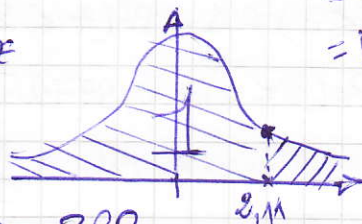
Ex 2 = $Pr(x \leq -2{,}11) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-2{,}11} e^{-\frac{x^2}{2}} dx$



$= Pr(x \geq 2{,}11) = 1 - Pr(x \leq 2{,}11)$

$= 1 - 0{,}98257$

$= \boxed{0{,}01743}$



Ex 3 = $Pr(-1 \leq x \leq 1) = ???$



$= 1 - [Pr(x \leq -1) + Pr(x \geq 1)]$

$Pr(x \leq -1) = 1 - Pr(x \leq 1)$

$Pr(x \geq 1) = 1 - Pr(x \leq 1)$

$P = 1 - [1 - Pr(x \leq 1) + 1 - Pr(x \leq 1)]$

$= 1 - 1 + Pr(x \leq 1) - 1 + Pr(x \leq 1)$

$= 2 Pr(x \leq 1) - 1 = 2 \times 0{,}8413 - 1 = \boxed{0{,}6826}$

Exercice ① = On mesure la profondeur x (en cm) d'enracinement d'une plante steppique dans le sol.

Pour 200 observations, on trouve une moyenne $\bar{x} = 32$ cm et un écart-type $\sigma = 7$ cm. On suppose que cette variante X suit une loi normale.

Donnez les probabilités suivantes :

1- $Pr(x > 32)$.

2- $Pr(32 \leq x \leq 32 + 2\sigma)$.

N^{bre}	profondeur	$(x_i - \bar{x})^2$
1	35	⋮
2	25	⋮
⋮	⋮	⋮
200	30	
		$\Sigma(x_i - \bar{x})^2$

1) $\bar{x} = \Sigma \frac{x_i}{200} = 32$ cm.

$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{200}} = 7$ cm.

on doit changer à la loi normale réduite ⇒

$Pr(x > 32) = Pr\left(\frac{x - 32}{7} > \frac{32 - 32}{7}\right) = Pr(t > 0)$

$= Pr(t \leq 0)$

$= 0{,}5$

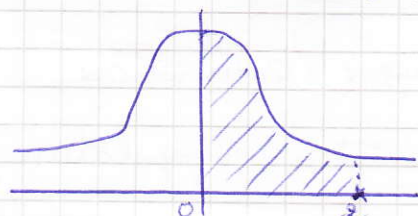
($\frac{x-32}{7} = t$: variable normale)

(Pour changer à une loi réduite : $\frac{x - \bar{x}}{\sigma}$; $\bar{x}$ → moyenne, σ ← écart-type)

2) $Pr(32 \leq x \leq 32 + 2\sigma)$

$= Pr(32 \leq x \leq 32 + 14)$

$= Pr(32 \leq x \leq 46)$



(transfert à loi normale) : $= Pr\left(\frac{32-32}{7} \leq \frac{X-32}{7}_t \leq \frac{46-32}{7}\right)$

$= Pr(0 \leq t \leq 2)$

$= 1 - [Pr(t \leq 0) + Pr(t \geq 2)]$

$= 1 - [Pr(t \leq 0) + 1 - Pr(t \leq 2)]$

$Pr(t \leq 2) - Pr(t \leq 0) = 0{,}97725 - 0{,}5 = \boxed{0{,}47725}$

Exercice 02 : Une coopérative agricole à fait l'acquisition d'1 nouvelle variété de blé. A la 1ère récolte, on a constaté que la longeur moyenne de 400 épis de blé est de 11,5 cm et un écart-type de 1,5 cm.

1- Quelle est la nature de la loi variante étudiée ? et quelle loi suit-elle Théoriquement ?

2- Quelle est la probabilité de trouver un épis :

a- de taille supérieure à 13 cm.

b- " " inférieure à 10 cm.

Donc : ($\bar{x} = 11{,}5$ cm), ($\sigma = 1{,}5$ cm).

Elle suit la loi normale.

1) La variable est continue, elle suit la loi normale.

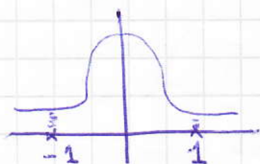
2) a) $Pr(X > 13) = Pr\left(\frac{X-11{,}5}{1{,}5}_t > \frac{13-11{,}5}{1{,}5}\right)$

$= Pr(t > 1) = 1 - Pr(t \leq 1) = 1 - 0{,}8413 = \boxed{0{,}1587}$

B) $Pr(X < 10) = Pr\left(\frac{X-11{,}5}{1{,}5}_t < \frac{10-11{,}5}{1{,}5}\right)$

$= Pr(t < -1)$

$= Pr(t > 1) = \boxed{0{,}1587}$



on suppose $\Longrightarrow$ $\boxed{Pr(X > 13) = Pr(X < 10)}$

(***) Variables continues :

1) loi normale.

2) " de Student.

3) " " χ^2 (Khi-deux).

4) " " Fisher

Prochaine séance : les Estimations

((on estime les paramètres d'1 population, moyenne ....))

le : 11.10.2017

Les Estimations

1 - **Introduction** : à défaut de pouvoir connaître les valeurs exactes des paramètres d'1 population, ce qui nécessiterait de mesurer tous les éléments (recensement), on cherche à les estimer.

Lorsque l'estimation est donnée par un seul nombre, l'estimation est ponctuelle.

Quand elle consiste à déterminer des nombres délimitant un intervalle ayant une ~~propa~~ probabilité donnée de renfermer le paramètre, on a une estimation par intervalle.

1) Estimation d'une moyenne =

1.1. Estimation ponctuelle :

Ayant affaire à 1 population dont les éléments présentent un caractère quantitatif, on veut estimer la moyenne inconnue d'une population μ (mu).

On calcule alors la moyenne de l'échantillon tiré au hasard :

$$\bar{x} = \frac{\sum x_i}{n}$$

N.B : $E(x) = \mu$ (la moyenne de la population = la moyenne de l'échantillon)

E : la moyenne des moyennes (Espérance mathématique)

x	$V(x) = \frac{\sum (x_i - E(x))}{30}$
$\bar{x}_1$	
⋮	
$\bar{x}_{30}$	
$\frac{\sum \bar{x}_i}{30}$	

1.2. Estimation par intervalle :

1er Cas = n grand ($n \geq 30$).

$$\bar{x} - z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

$\bar{x}$: moyenne de l'échantillon.

σ : Ecart-type " " .

n : effectif " " .

α : Risque d'erreur ($\alpha = 5\%$).

Z : → Déterminer dans la table de la loi normale.

Exemple :

$n = 100$; $\bar{x} = 50$; $\sigma^2 = 400$.

① Calculer l'intervalle de confiance de la moyenne de la population $\alpha = 5\%$.

$$\bar{x} - Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}.$$

table =

	0,00 - - - - - 0,06
0,0	
0,1	
⋮	
1,9	- - - 0,975

$$Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0,05}{2}} = Z_{0,975} = 1,9 + 0,06 = \boxed{1,96}.$$

Donc : $50 - 1,96 \times \frac{20}{10} \leq \mu \leq 50 + 1,96 \times \frac{20}{10}$

$46,08 \leq \mu \leq 53,92 \Rightarrow \mu = 50 \pm 3,92$

donc = $\mu \in [46,08 , 53,92]$.

⇓

(95% de chance pour que la moyenne appartienne à cet intervalle.)

2) Calculer cette intervalle pour $\alpha = 1\%$.

$$Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0,01}{2}} = Z_{1-0,005} = Z_{0,995} = 2,5 + 0,08 = \boxed{2,58}$$

donc : $50 - 2,58 \times \frac{20}{10} \leq \mu \leq 50 + 2,58 \times \frac{20}{10}$

$50 - 5,16 \leq \mu \leq 50 + 5,16 \Rightarrow 44,84 \leq \mu \leq 55,16$

$\Rightarrow \mu = 50 \pm 5,16$

donc $\mu \in [44,84 , 55,16]$.

le : 16.10.2017 =

2ème cas : n petit ($n < 30$).

$$\bar{x} - t_{1-\frac{\alpha}{2}} \times \frac{\sigma_{ech}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\frac{\alpha}{2}} \times \frac{\sigma_{ech}}{\sqrt{n}}$$

L'intervalle de confiance de la moyenne est donné :

$$\bar{x} - t_{1-\frac{\alpha}{2}} \times \frac{\sigma_{ech}}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{1-\frac{\alpha}{2}} \times \frac{\sigma_{ech}}{\sqrt{n-1}}$$

$t_{1-\frac{\alpha}{2}}$ Table suivante : ???

Exemple : un dosage de sucre dans une solution effectué sur 8 prélèvements provenant d'1 même fabrication a donné les résultats suivants (exprimés en g/l) :
19,5 - 19,8 - 20,2 - 20,2 - 20,3 - 20,5 - 20,8 - 19,7.

1) Calculez la moyenne et l'écart-type de cette distribution ?
2) " l'intervalle de confiance de la moyenne au risque $\alpha = 5\%$?.

x_i	n_i	$n_i x_i$	$n_i(x_i - \bar{x})^2$
19,5	1	19,5	0,39
19,8	1	19,8	0,10
20,2	2	40,4	⋮
20,3	1	20,3	
20,5	1	20,5	
20,8	1	20,8	
19,7	1	19,7	
	$n = 8$	$\sum n_i x_i$	$\sum n_i(x_i - \bar{x})^2$

1) $\sigma^2_{éch} = \frac{\sum n_i (x_i - \bar{x})^2}{n}$

$\bar{x} = \frac{\sum n_i x_i}{n} = 20,125$ g/l

$\sigma^2 = \sqrt{0,22} = 0,46$ g/l.

2) $\bar{x} - t_{1-\frac{\alpha}{2}} \times \frac{\sigma_{éch}}{\sqrt{n-1}}$

$\nu = n - 1 = 8 - 1 = 7$

$t_{1-\frac{\alpha}{2}} = t_{1-\frac{0,05}{2}} = 0,975$

$20,12 - 2,36 \times \frac{0,46}{\sqrt{7}} \leq \mu \leq 20,12 + 2,36 \times \frac{0,46}{\sqrt{7}}$

$\Rightarrow 19,71 \leq \mu \leq 20,53 \Rightarrow \mu \in [19,71 , 20,53]$.

2) Estimation d'une proportion :

2.1. Estimation ponctuelle : Dans la pratique, on dispose d'1 seul échantillon, on calculera alors dans l'échantillon la proportion P et on écrit $\boxed{\Pi = p}$

120 agriculteurs → 80 pour, → 40 contre $\Rightarrow P = \frac{80}{120} = 0,66 \Rightarrow 66\% \Rightarrow \Pi = 66\%$

2.2. Estimation par intervalle : l'intervalle de confiance d'1 proportion s'écrit : $P - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}} \leq \mu \leq P + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}}$

P : proportion dans l'échantillon.
n : effectif dans "
Π : proportion estimée de la population.
$Z_{1-\frac{\alpha}{2}}$: Table du loi normale.

Exemple : on considère un échantillon de 169 brebis de race Ile de France. Ces brebis ont été mise lutte, on a 108 brebis pleines (fécondes).
- Donnez l'intervalle de confiance à 95% du taux de fertilité de cette race,

$P = \frac{108}{169} = 0,64$; $Z_{1-\frac{\alpha}{2}} = Z_{0,975} = 1,96$.

$0,64 - 1,96 \times \sqrt{\frac{0,64(1-0,64)}{169}} \leq \Pi \leq 0,64 + 1,96 \times \sqrt{\frac{0,64(1-0,64)}{169}}$

$\Rightarrow 0,56 \leq \Pi \leq 0,71 \Rightarrow 56\% \leq \Pi \leq 71\%$.

le : 18.10.2017 :

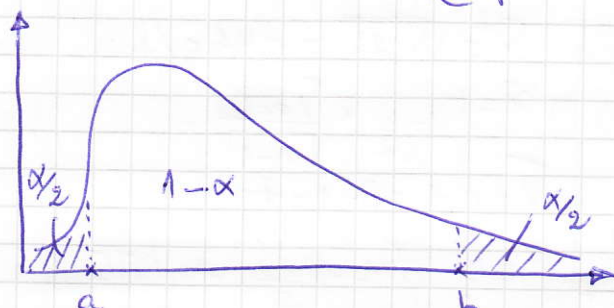
3. Estimation d'une variance par intervalle : loi de χ^2 (Khi-deux) est une loi de probabilité continue et dépend du N° du degré de liberté

* Si : $n \leq 30$.

pour α donné, on détermine les nombres a et b tel que

$Pr(Y \leq a) = \frac{\alpha}{2}$ et $Pr(Y \geq b) = \frac{\alpha}{2}$.

⇒ on utilise la table du χ^2 pour déterminer



L'intervalle de confiance de la variance $\frac{(n-1)\sigma^2_{échan}}{b} \leq \sigma^2 \leq \frac{(n-1)\sigma^2_{éch}}{a}$

Exercice : on a mesuré le poids de raisin par souche ; sur 10 souches prises au hasard dans une vigne, on a obtenu les résultats suivants exprimés en kg : 2,7 − 3,2 − 3,6 − 4,1 − 4,3 − 4,7 − 5,4 − 5,9 − 6,5 − 6,9.

a) Donnez un intervalle de confiance de la moyenne de la population $\alpha = 5\%$.

b) Donnez pour $\alpha = 5\%$, un intervalle de confiance de la vraie variance de la population.

a)

x_i	$(x_i - \bar{x})^2$
2,7	4,12
3,2	2,34
3,6	1,27
4,1	0,39
4,3	0,18
4,7	0,0009
5,4	0,44
5,9	1,37
6,5	3,13
6,9	4,71
$\Sigma x_i =$ 47,3	$\Sigma(x_i - \bar{x})^2 =$ 17,95

* la moyenne : $\bar{x} = \frac{\Sigma x_i}{n} = \frac{47,3}{10} = 4,73$ kg

* la variance $\sigma^2_{éch} = \frac{\Sigma(x_i - \bar{x})^2}{n}$

$= \frac{17,95}{10} = 1,8$ kg²

* Ecart type $\sigma = \sqrt{1,8 \text{ kg}^2} = 1,34$ kg.

* Intervalle de confiance de la moyenne

$\bar{x} - t_{1-\frac{\alpha}{2}} \times \frac{\sigma_{éch}}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{1-\frac{\alpha}{2}} \times \frac{\sigma_{éch}}{\sqrt{n-1}}$

$t_{1-\frac{0,05}{2}} = 0,975$.

ddl $= n-1 = 10-1 = 9$.

$t = 2,26$ (table)

$$4,73 - 2,26 \times \frac{1,8}{\sqrt{9}} \leq \mu \leq 4,73 + 2,26 \times \frac{1,8}{\sqrt{9}}$$

$$3,374 \leq \mu \leq 6,086.$$

$\Rightarrow$ Donc : $\mu \in [3,374 ; 6,086]$.

b) Intervalle de confiance de la variance

$$\frac{(n-1)\,\sigma^2_{ech}}{b} \leq \sigma^2 \leq \frac{(n-1)\,\sigma^2_{ech}}{a}$$

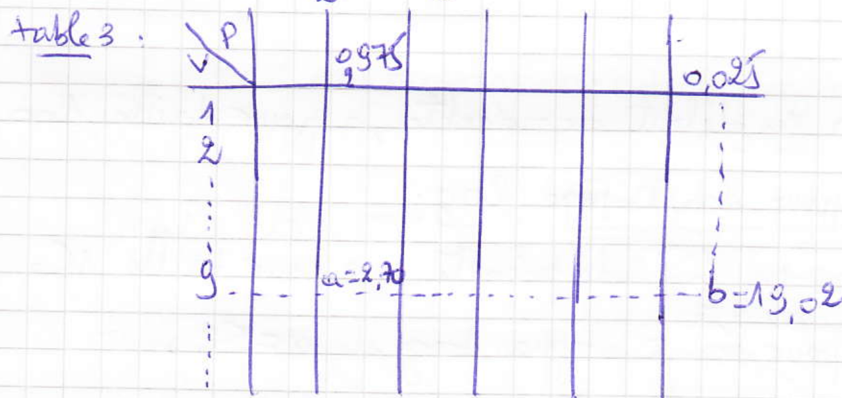
$(Pr(y \leq b) \leq 1 \leq Pr(y \geq a))$

$Pr(y \geq a) = 1 - \frac{\alpha}{2} = 1 - \frac{0,05}{2} = 0,975$

$Pr(y \geq b) = \frac{\alpha}{2} = \frac{0,05}{2} = 0,025$

$\frac{\alpha}{2}$ — a — b

table 3 :



N.B : moyenne > 30 $\Rightarrow$ loi normale ; " < 30 $\Rightarrow$ " Student ; Variance $\Rightarrow \chi^2$

Donc : $\frac{(10-1)\times 1,8}{19,02} \leq \sigma^2 \leq \frac{(10-1)\times 1,8}{2,7}$

$$0,85 \leq \sigma^2 \leq 6$$

le : 23.10.2017 =

Tests d'hypothèses

$\Rightarrow$ I) Test de comparaison de 2 moyennes = 1) n_1 et $n_2 \geq 30$, $H_0 : \mu_1 = \mu_2$

Dans 2 populations P_1 et P_2, on étudie une variable aléatoire x,

on note : μ_1 et σ_1 désignent la moyenne et l'écart type de la population P_1

μ_2 et σ_2 " " " " " " " P_2

(Ces moyennes et les écart-types sont inconnus).

Dans P_1, on extrait un échantillon E_1, de taille n_1, pour lequel on calcule la moyenne $\bar{x}_1$ et l'écart-type σech_1.

Dans P_2 on extrait un échantillon E_2, de taille n_2, pour lequel on calcule la moyenne $\bar{x}_2$ et l'écart-type σech_2.

{° hypothèse à tester $H_0 : \mu_1 = \mu_2$

- calcul de l'écart-moyen : $\varepsilon = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{\sigma^2_{ech}}{n_1} + \dfrac{\sigma^2_{ech}}{n_2}}}$

- $Z_\alpha \rightarrow$ Table de la loi normale.
  pour α fixe

Si $\varepsilon < U_\alpha \longrightarrow H_0$ ~~rejetée~~ acceptée.

$\varepsilon \geqslant U_\alpha \longrightarrow H_0$ ~~acceptée~~ rejetée.

Exercice ① = On a prélevé 2 échantillons de pomme de terre pour étudier les poids.

- le 1er échantillon en début de récolte, a pour taille 100, pour moyenne 120g et pour écart-type 20g.
- le 2ème échantillon en fin de récolte, a pour taille 150, pour moyenne 150g et pour écart-type 10g. $\alpha = 5\%$.

$\longrightarrow$ La différence entre les 2 poids moyens à ces 2 époques différent est-elle significative ?

Solution = H_0 : les poids ne sont pas significativement différents aux 2 époques.

$$\varepsilon = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma^2_{ech\,1}}{n_1} + \frac{\sigma^2_{ech\,2}}{n_2}}} = \frac{|120-150|}{\sqrt{\frac{400}{100} + \frac{100}{150}}} = \frac{|-30|}{\sqrt{4+0{,}66}} = \frac{30}{\sqrt{4{,}66}} = 13{,}88$$

$Z_{1-\alpha/2} = Z_{0,975} \longrightarrow$ loi normale : $Z_{1-\frac{\alpha}{2}} = 1{,}96$.

$\varepsilon > Z_{1-\frac{\alpha}{2}} \longrightarrow H_0$ rejetée.

$\Rightarrow$ les poids aux 2 époques différents d'une manière significative

Exercice ② = Dans 2 parcelles de blé A et B, on a compté le nombre de plants en mètre linéaire sur plusieurs placettes en vue de déterminer le peuplement. Les résultats sont les suivants :

Données	Parcelle A	Parcelle B
Nombre de Comptage	32	40
Nbre moyen de plants / mètre linéaire	51,8	49,3
Variance	25,45	22,7

— Les peuplements de ces 2 parcelles sont-ils différents au seuil $\alpha = 5\%$?

Ho : les peuplements des 2 parcelles sont identiques

$$\varepsilon = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma^2_{ech_1}}{n_1} + \frac{\sigma^2_{ech_2}}{n_2}}} = \frac{|51,8 - 49,3|}{\sqrt{\frac{25,45}{32} + \frac{22,7}{40}}} = \frac{2,5}{\sqrt{0,79 + 0,56}} = \frac{2,5}{\sqrt{1,35}} = 2,15$$

$z_{1-\frac{\alpha}{2}} = z_{0,975}$ → loi normale : $z_{1-\frac{\alpha}{2}} = 1,96$.

$\varepsilon > 1,96$ → rejetée

⇒ Donc les peuplements des 2 parcelles sont différents.
(" " différent d'1 manière significative).

le : 25.10.2017

Résumé = Test de comparaison de 2 moyennes.

1) n_1 et $n_2 \geq 30$.

$H_0 : \mu_1 = \mu_2$ → $\varepsilon = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma^2_{ech_1}}{n_1} + \frac{\sigma^2_{ech_2}}{n_2}}}$

table loi normale pour α fixé → z_{th}

$\varepsilon < z_{th}$ → H_0 acceptée

$\geq z_{th}$ → H_0 rejetée

2) n_1 et $n_2 < 30$, $H_0 : \mu_1 = \mu_2$.

au lieu de l'écart réduit (applicable pour les grands échantillons), on utilise le critère de Student : $t = \frac{|\bar{x}_1 - \bar{x}_2|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

s : Écart type commun aux 2 échantillons

$$s^2 = \frac{(n_1 - 1)\sigma^2_{ech_1} + (n_2 - 1)\sigma^2_{ech_2}}{n_1 + n_2 - 2}$$

Table de Student pour déterminer un t théorique — α fixé, ddl $= n_1 + n_2 - 2$ (ν)

Si : $t < t_{th}$ → H_0 : retenue.

$t \geq t_{th}$ → H_0 : rejetée.

Exercice : Pour déterminer le poids moyen d'épis de blé appartenant à 02 variétés, on a procédé à 10 pesées pour chaque variété

les moyennes obtenues : $\bar{x}_1 = 170,7$ cg ; $\bar{x}_2 = 168,5$ cg.

Les variances : $\sigma^2_{ech_1} = 432,9$ cg^2 ; $\sigma^2_{ech_2} = 182,7$ cg^2.

- la ≠ des moyennes est-elle significative au risque $\alpha = 5\%$?

$$S_{*}^{2} = \frac{(10-1)\times 432,9 + (10-1)\times 182,7}{10+10-2} = \frac{9\times(432,9+182,7)}{18} = \frac{615,6}{2} = 307,8.$$

$s^2 = 307,8 \longrightarrow s = \sqrt{307,8} = 17,54.$

donc : $t = \frac{|170,7 - 168,5|}{17,54\sqrt{\frac{1}{10}+\frac{1}{10}}} = 0,28$

ν	... $t_{0,975}$
18	... 2,10

on a : $\alpha = 5\% \rightarrow t_{1-\alpha/2} = t_{0,975}$

$\nu = n_1 + n_2 - 2 = 10 + 10 - 2 = 18$

$t_{th} = 2,10$

$\Rightarrow t < t_{th} \rightarrow H_0$ retenue, Il n'y a de différence significative entre les 2 moyennes.

II) Comparaison d'une moyenne avec une valeur théorique.

1) $n \geq 30$: $H_0 : \mu = \mu_0$

$$Z = \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}$$

table loi normale $\longrightarrow Z_{th}$ pour α fixé

Si $Z < Z_{th} \longrightarrow H_0$ acceptée.

Si $Z \geq Z_{th} \longrightarrow H_0$ rejetée.

Exercice : Les spécifications d'1 certain médicament indiquent que chaque comprimé doit contenir 2,5g de substance active. 100 comprimés choisis au hasard dans la production, puis analysés contiennent en moyenne 2,6g de substance active avec un écart-type de 0,4g.

- Peut-on dire que le médicament respecte les spécifications ? $\alpha = 5\%$

Solution : $H_0 : \mu = 2,5g$.

$$Z = \frac{|\bar{x} - 2,5|}{\sigma/\sqrt{n}} = \frac{|2,6 - 2,5|}{0,4\sqrt{100}} = 0,025.$$

Z_{th} pour $\alpha = 5\%$; $Z_{1-\alpha/2} = Z_{0,975} = 1,96$.

$\Rightarrow Z < Z_{th} \longrightarrow H_0$ retenue (le médicament respecte les spécification).

$0,025 < 1,96$

le : 30.10.2017 =

III) Test de comparaison de plus de 2 moyennes : Analyse de la variance

1. Introduction =

On dispose de k échantillon indépendants $E_1, E_2 \cdots E_k$ extraits de k populations $P_1, \ldots P_k$ supposées Gaussiennes et de même variance σ^2.
Les moyennes respectives des populations sont notés $\mu_1, \mu_2 \cdots \mu_k$.

L'analyse de la variance permet de comparer globalement les moyennes des populations.

L'hypothèse nulle à ~~comparer~~ vérifier est donc : $H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

2. Eléments d'analyse =

Pour chaque échantillon E_i, de taille n_i, on calcule la moyenne $\bar{x}_i$ et la variance $\sigma^2_{ech_i}$.

En plus de ces caractéristiques, on doit déterminer =

2.1. La variance factorielle : (V_f)

Elle caractérise la dispersion d'un échantillon à un autre, c'est à dire la variance due à l'influence du facteur étudié.

$$V_f = \frac{1}{k-1} \sum (\bar{x}_i - \bar{x})^2 = \frac{\left[n_1 (\bar{x}_1)^2 + n_2 (\bar{x}_2)^2 + \ldots + n_k (\bar{x}_k)^2 - n(\bar{x})^2\right]}{k-1}$$

$\bar{x}$: moyenne générale. $\bar{x} = \frac{\sum \bar{x}_i}{n}$, $n = \sum n_i$.

k : Nbre d'échantillon.

2.2. La variance résiduelle (V_r) :

$$V_r = \frac{1}{n-k}\left[\sum (x_{i_1} - \bar{x}_1)^2 + \sum (x_{i_2} - \bar{x}_2)^2 + \ldots \sum (x_{i_k} - \bar{x}_k)^2\right]$$

$$V_r = \frac{1}{n-k}\left[\sum x_i^2 - (n_1 \bar{x}_1^2 + n_2 \bar{x}_2^2 + \ldots + n_k \bar{x}_k^2)\right]$$

3. Degré de liberté ddl : ν_1 et ν_2 (nu) :

ν_1 est localisé dans la 1ère ligne de la table de Fisher.

ν_2 " " " " 2ème colonne.

$\nu_1 = k-1$, $\nu_2 = n-k$

L'intersection de V_1 et V_2 dans la table de fisher, selon le risque α choisi donne directement la valeur théorique de F_α fisher.

- Si $\frac{V_f}{V_r} < F_\alpha \longrightarrow H_0$ retenue.
- Si $\frac{V_f}{V_r} \geq F_\alpha \longrightarrow H_0$ rejetée.

Ex : On a relevé les taux suivants de glycemie en g/l de plasma, chez 03 groupes de sujets. $H_0 : \mu_A = \mu_B = \mu_C$

Groupe A		Groupe B		Groupe C	
x_{iA}	x_{iA}^2	x_{iB}	x_{iB}^2	x_{iC}	x_{iC}^2
0,80	0,64	1,30	1,69	1,15	1,32
1,00	1,00	1,30	1,69	1,20	1,44
1,00	1,00	1,60	2,56	1,96	3,84
1,18	1,3924	0,88	0,77	1,98	3,92
0,87	0,7569	0,95	0,90	1,13	1,27
0,82	0,6724	1,35	1,82	1,29	1,66
0,95	0,9025	1,21	1,46	1,25	1,56
1,05	1,1025	1,60	2,56	0,98	0,96
0,73	0,5329	1,70	2,89	1,81	3,27
		0,75	0,56	0,85	0,72
		1,50	2,25	1,60	2,56
				1,69	2,85
$n_A = 9$ $\sum x_{iA} = 8,4$ $\bar{x}_A = 0,93$ $\bar{x}_A^2 = 0,8649$	$\sum x_{iA}^2 = 9,8472$ 7,99	$n_B = 11$ $\sum x_{iB} = 14,14$ $\bar{x}_B = 1,285$ $\bar{x}_B^2 = 1,651$	$\sum x_{iB}^2 = 19,15$	$n_C = 12$ $\sum x_{iC} = 16,89$ $\bar{x}_C = 1,407$ $\bar{x}_C^2 = 1,979$	$\sum x_{iC}^2 = 25,37$ 25,4

Le : 06.11.2017 =

$H_0 : \mu_A = \mu_B = \mu_C$

Calcul de la variance factorielle :

$\bar{x} = \frac{\sum \bar{x}_i}{N} = \frac{\bar{x}_A + \bar{x}_B + \bar{x}_C}{32} = 1,24$

$(N = n_A + n_B + n_C = 9 + 11 + 12 = 32)$.

$\sum x_i^2 = \sum x_{iA}^2 + \sum x_{iB}^2 + \sum x_{iC}^2$
$= 52,55$.

$$V_f = \frac{1}{k-1} \sum n_i (\bar{x}_i - \bar{x})^2$$

$$V_f = \frac{1}{k-1} * [n_A(\bar{x}_A)^2 + n_B(\bar{x}_B)^2 + n_C(\bar{x}_C)^2 - N(\bar{x})^2]$$

$$V_f = \frac{1}{3-1} \times 9(0,93)^2 + 11(1,28)^2 + 12(1,41)^2 - 32(1,24)^2$$

$$\Rightarrow \boxed{V_f = 0,29}$$

Calcul de la variance résiduelle :

$$V_r = \frac{1}{N-k} \left[\sum x_i^2 - (n_A \bar{x}_A^2 + n_B \bar{x}_B^2 + n_C \bar{x}_C^2)\right]$$

$$V_r = \frac{1}{32-3} \left[52,55 - (9 \times (0,93)^2 + 11(1,28)^2 + 12(1,41)^2)\right]$$

$$\Rightarrow \boxed{V_r = 0,12}$$

$$\frac{V_f}{V_r} = \frac{0,29}{0,12} = 2,42$$

$$\Rightarrow F_{th}(\nu_1, \nu_2, \alpha) = F(2, 29, 5\%) = 3,33 \Rightarrow \text{Tableau}$$

($2 = k-1$; $29 = N-3$)

	1	2
⋮		
28		3,34
29		3,33
30		3,32

donc : $\frac{V_f}{V_r} < F_{th} \Rightarrow H_0$ retenue

IV) Test de conformité :

1. Comparaison d'une répartition observée à une répartition théorique : test du χ^2 : (Khi-deux).

La répartition théorique ayant été choisie, il s'agit de savoir si la répartition expérimentale est conforme à la répartition théorique choisie. La vérification de cette conformité est faite à l'aide du test du χ^2

- Définition du χ^2 :

$$\chi^2 = \frac{\sum (O_i - C_i)^2}{C_i}$$

O_i : Effectif observé ou expérimental

C_i : " Calculé ou théorique.

H_0 : La répartition expérimentale est conforme à la répartition théorique

- on calcule $\chi^2 = \frac{\sum (O_i - C_i)^2}{C_i}$.
- on détermine la valeur χ^2 théorique dans la table du χ^2 en fonction

de α et de $ddl = k-1$ (k = nbre de classes).

- $\chi^2 < \chi^2_{théorique}$, H_0 : retenue.
- $\chi^2 \geq \chi^2_{théorique}$, H_0 : rejetée.

Remarque : le test du χ^2 ne peut être utilisé que si les effectifs théoriques sont suffisament grands ; $c_i \geq 5$.

Exercice : on a croisé 2 lignées pures présentant les caractères A et B. En F_2, on observe 4 types de plantes de phénotypes AB, Ab, aB et ab dont les effectifs sont respectivements : 100, 28, 20 et 12.

La loi de Mendel prévoit les proportions : 9/16 - 3/16 - 3/16 et 1/16.

- Peut-on considérer avec une confiance de 0,95 que les effectifs observés sont conformes à la loi de Mendel ???

⇒ - H_0 : la répartition expérimentale est conforme à la loi de Mendel

- $\chi^2 = \sum \frac{(o_i - c_i)^2}{c_i}$.

$c_1 = 9/16 \times 160 = 90$
$c_2 = 3/16 \times 160 = 30$ ...

Phénotypes	AB	Ab	ab	ab	total
Proportions théoriques	9/16	3/16	3/16	1/16	1
effectifs calculés c_i	90	30	30	10	160
effectifs observés o_i	100	28	20	12	160

$$\chi^2 = \frac{(100-90)^2}{90} + \frac{(28-30)^2}{30} + \frac{(20-30)^2}{30} + \frac{(12-10)^2}{10} = 4,97.$$

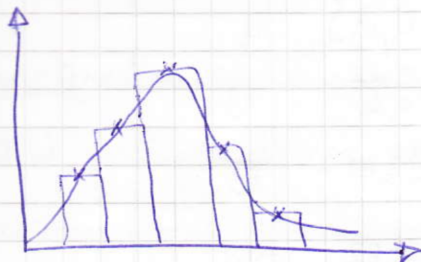
$\chi^2_{théorique}$ < $ddl = k-1 = 4-1$, $\alpha = 5\%$ ⇒ $\chi^2_{théorique} = 7,81$.

⇒ $\chi^2 < \chi^2_{th}$ → H_0 retenue.

c'est à dire la répartition expérimentale est conforme à la loi de Mendel.

le : 08.11.2017 =

→ Test de normalité :



⟹ Courbe de la loi normale



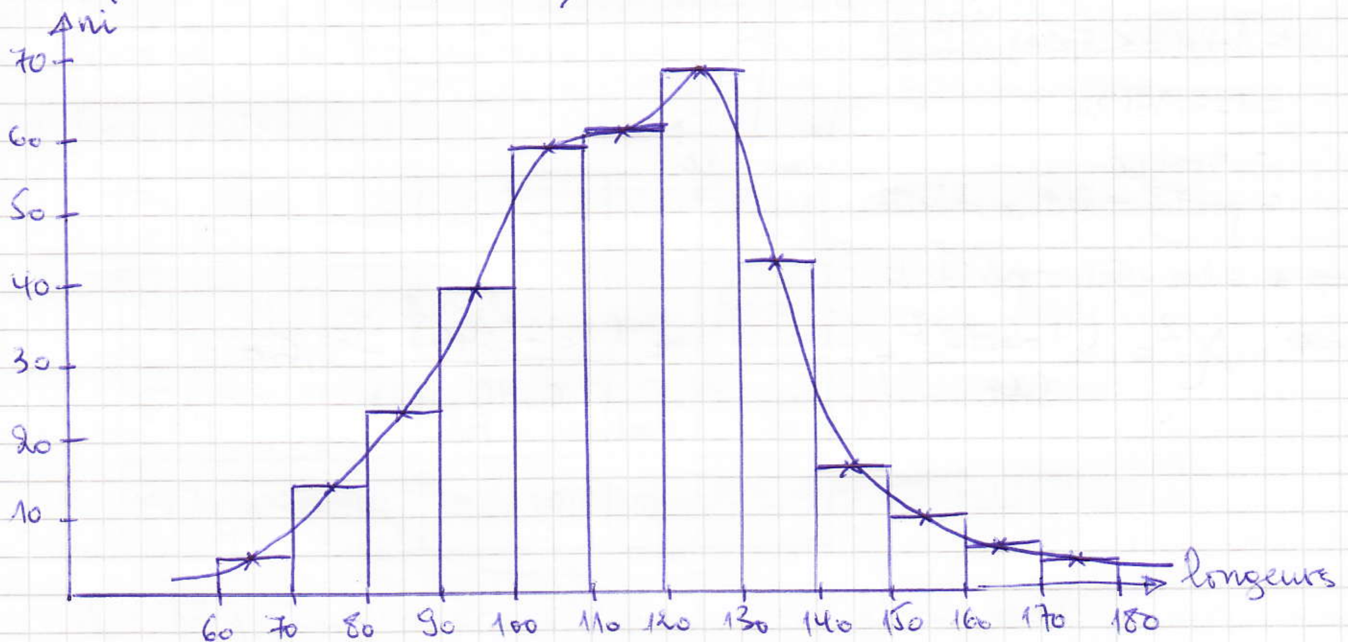
H_0 : les longueurs des épis de Maïs suit une loi normale.

Une 1ère indication peut être donnée par le tracé d'un histogramme selon que la représentation suggère ou non la forme de la distribution normale.

Exercice : D'une population d'épis de maïs, on extrait d'une manière aléatoire un échantillon d'effectif 345 et l'on a mesuré les longueurs en mm.

Longeurs (mm)		60-70	70-80	80-90	90-100	100-110	110-120	120-130	130-140	140-150	150-160	160-170	170-180
Effectifs (n_i)		4	13	23	40	58	60	67	44	16	10	6	4

Les caractéristiques statistiques de la distribution des longeurs des épis de Maïs sont : $\bar{x} = 115,52$ mm ; $\sigma_{ech} = 21,35$ mm.



– Calcul du χ^2 : $\chi^2 = \frac{\sum (O_i - C_i)^2}{C_i}$

$\frac{60-115{,}58}{21{,}35}$

Classes	$z_i = \frac{x_i - \bar{x}}{\sigma}$	$\Pi(z_i) = P_r(z \leq z_i)$	$P_i = \Pi(z_{i+1}) - \Pi(z_i)$	$C_i = P_i \times n$ (345)	O_i
$-\infty$ - 60			0,00466	1,61 }	0
	−2,60	0,00466			
60 - 70 (*) −2,60 −2,13			0,01193	4,11 }	4
	−2,13	0,01652			
70 - 80			0,03155	10,88	13
	−1,66	0,48846			
80 - 90			0,06785	23,41	23
	−1,19	0,11600			
90 - 100			0,11980	41,33	40
	−0,72	0,23580			
100 - 110			0,16550	57,09	58
	−0,25	0,40130			
110 - 120			0,18190	62,75	60
	+ 0,21	0,58320			
120 - 130			0,16540	57,66	67
	0,67	0,74860			
130 - 140			0,12430	42,88	44
	1,14	0,87290			
140 - 150			0,07340	25,32	16
	1,61	0,94630			
150 - 160			0,03494	12,05	10
	2,08	0,98124			
160 - 170			0,013370	4,61 }	6 }
	2,55	0,99461			
170 - 180			0,004080	1,40 } 6,46	4 } 10
	3,01	0,99869			
180 - $+\infty$			0,001300	0,45 }	0 }

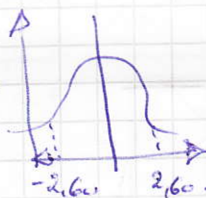
$\Rightarrow \Pi(z_i) = Pr(Z \leq -2{,}60)$

$= 1 - Pr(Z \leq 2{,}60)$ tableau loi normale

$= 1 - 0{,}99534$

$= 0{,}00466$.

si positif → lecture directe



⇒ * valeur − valeur précédente

$$\Rightarrow \chi^2 = \frac{(4-5{,}72)^2}{5{,}72} + \cdots + \frac{(10-6{,}46)^2}{6{,}46} = 8{,}60$$

Le : 13.11.2017 =

Test d'indépendance

1) Caractères qualitatifs =

Le test d'indépendance entre 2 caractères qualitatifs se fait selon les étapes suivantes :

1. Formuler l'hypothèse nulle : H_0 : les 2 caractères qualitatifs sont indépendants.
2. Calcul du X^2 : $X^2 = \frac{\sum (o_i - c_i)^2}{c_i}$.

On calculera à partir du tableau de contingence pour chaque case, l'effectif calculé : C_{ik} = total des effectifs de la ligne i x le total des effectifs de la colonne k divisé par l'effectif total.

3. Pour α fixé, et un ddl = (l-1)(c-1), on détermine la valeur de X^2 théorique dans la table du X^2.
4. Décision : si $X^2 \geq X^2_{th}$ → H_0 rejetée.
   si $X^2 < X^2_{th}$ → H_0 acceptée.

Exercice : Afin de déterminer s'il y a indépendance entre le groupe sanguin et le sexe, on a examiné 976 individus prélevé au hasard. Les répartitions observées se répartissent dans le tableau suivant :

Sexe \ G.S	AB	A	O	B	total
Masculin	25	215	200	60	500
Féminin	15	207	194	60	476
Total	40	422	394	120	976

(annotations : × total / total)

H_0 : Il y a indépendance entre le G.S et le sexe.

calcul : $X^2 = \frac{\sum (O_i - c_i)^2}{c_i}$

G.S / Sexe	AB	A	B	O	total
M	20,49	216,19	201,84	61,48	500
F	19,51	205,81	192,16	58,52	476
Total	40	422	394	120	376

$$\Rightarrow \chi^2 = \frac{(25-20,49)^2}{20,49} + \frac{(215-216,19)^2}{216,19} + \cdots + \frac{(60-58,52)^2}{58,52} = 2,15.$$

$\alpha = 5\%$

$ddl = (2-1)(4-1) = 3$ } $\chi^2_{\text{théorique}}$ (table 3 χ^2).

	0,05
1	
2	
3	7,81

$\Rightarrow \chi^2_{\text{théo}} = 7,81$.

$\Rightarrow \chi^2 < \chi^2_{\text{théo}} \longrightarrow H_0$: retenue.

C'est à dire, il y a indépendance entre le G.S et le sexe.

2) Caractères Quantitatifs :

1- Regression et corrélation :

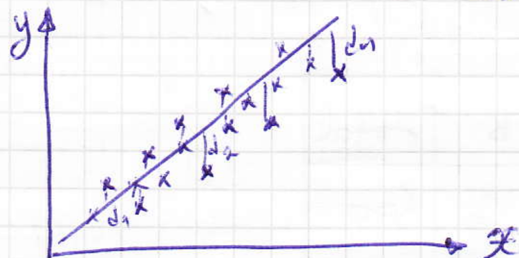
Supposons que l'étude porte sur 2 caractères quantitatifs, il s'agit de savoir s'il existe une liaison ou une corrélation (= liaison) entre ces 02 caractères.

ex : Poids et la taille

Poids → y ; taille → x

y : Prendra les valeurs $y_1, y_2, \ldots y_n$.

x : " " " $x_1, x_2, \ldots x_n$.



Relation linéaire : $y = a x + b$ (a ? b ?)

La meilleure droite est la droite dont les distances des différents points par rapport à la droite est minimale.

$\sum d_i = \text{Min}$, c'est la méthode des moindres carrés (المربعات الصغرى)

$$a = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{V(x)} = \frac{Cov(x,y)}{V(x)}$$

$\bar{y} = a\bar{x} + b \Rightarrow b = \bar{y} - a\bar{x}$

L'intensité de la liaison → Coefficient de corrélation.

$$Cov(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$V(x) = \sum (x_i - \bar{x})^2$$

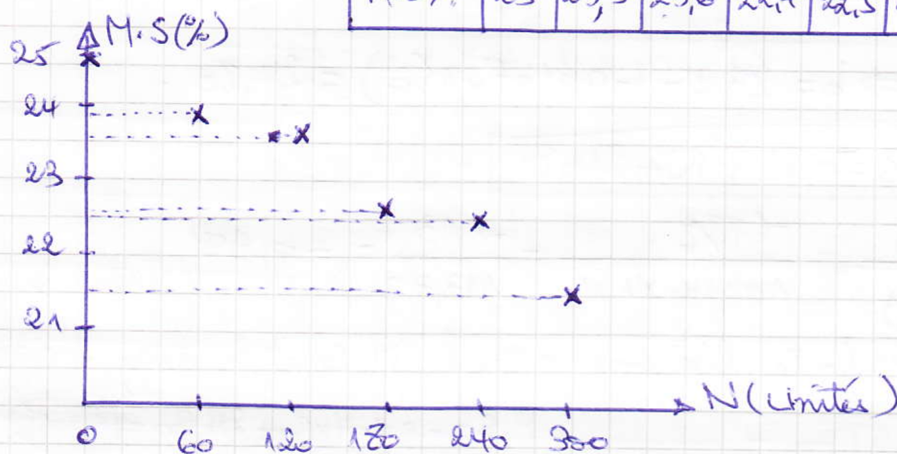
L'intensité de la liaison → coefficient de corrélation :

$$r = \frac{Cov(x,y)}{\sigma_x * \sigma_y} \quad , \quad -1 \leq r \leq 1$$



Exercice : Dans une expérimentation, on a montré que la teneur en matière sèche des tubercules de pomme de terre varient en fonction de la dose d'azote.

N (unités)	0	60	120	180	240	300
M.S %	25	23,9	23,6	22,7	22,5	21,5



$M.S = \underline{a}\,N + \underline{b}$

N x_i	M.S y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
0	25	-150	1,8	-270	22500	3,24
60	23,9	-90	0,7	-63	8100	0,49
120	23,6	-30	0,4	-12	900	0,16
180	22,7	30	-0,5	-15	900	0,25
240	22,5	90	-0,7	-63	8100	0,49
300	21,5	150	-1,7	-255	22500	2,89
$\sum x_i = 900$	$\sum y_i = 139,2$			$\sum (x_i - \bar{x})(y_i - \bar{y}) = -678$	$\sum (x_i - \bar{x})^2 = 63000$	$\sum (y_i - \bar{y})^2 = 7,32$

$\bar{x} = \frac{\sum x_i}{n} = 150 \left(\frac{900}{6}\right)$ u

$\bar{y} = \frac{\sum y_i}{n} = 23,2 \left(\frac{139,2}{6}\right)$ %

Le : 15.11.2017 =

$\bar{x} = \frac{\sum x_i}{n} = \frac{900}{6} = 150$ u.

$\bar{y} = \frac{\sum y_i}{n} = \frac{139,2}{6} = 23,2\ \%$.

$v(x) = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{63000}{6} = 10500\ u^2$

$\sigma(x) = \sqrt{v(x)} = \sqrt{10500} = 102,46$ u.

$v(y) = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{7,52}{6} = 1,25\ \%^2$

$\sigma(y) = \sqrt{v(y)} = \sqrt{1,25} = 1,11\ \%$

$a = \frac{Cov(x)}{v(x)} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{-678}{63000} = \boxed{-0,0107}$

$\left(\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} = Cov(x). \quad \frac{\sum (x_i - \bar{x})^2}{n} = v(x) \right)$ N.B

$\bar{y} = a\bar{x} + b$

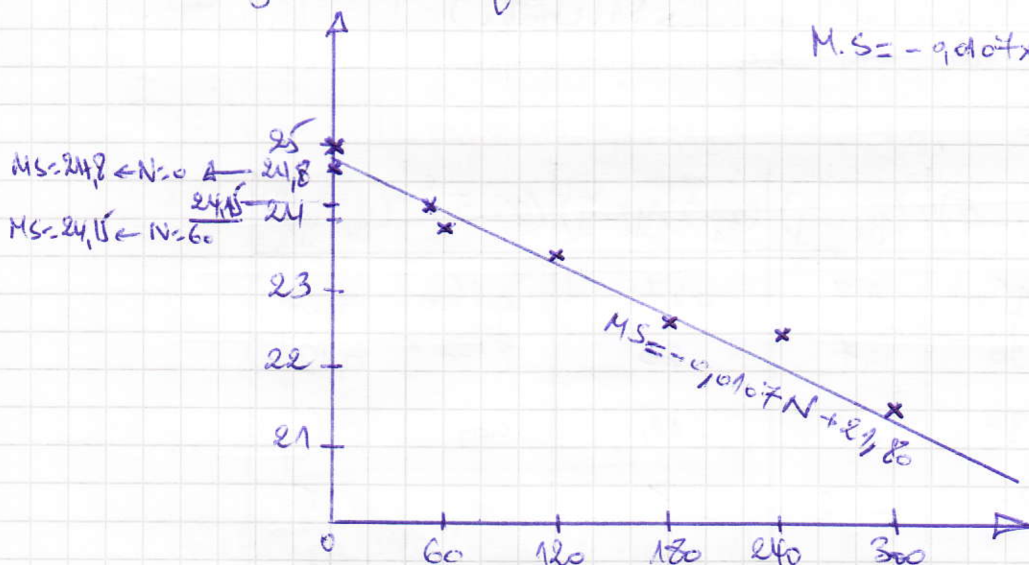
$b = \bar{y} - a\bar{x} \Rightarrow b = 23,2 - (-0,0107 \times 150) = 24,80$.

Donc $\Rightarrow$ $\boxed{MS = -0,0107\,N + 24,80}$

$r = \frac{Cov(x,y)}{\sigma(x) \times \sigma(y)} = \frac{-678/6}{102,46 \times 1,11} = \frac{-113}{113,73} = -0,99$

il y a une très forte liaison entre la M.S et la doze d'azote.

M.S = $-0,0107 \times 60 + 24,80 = 24,15$



expérimenter le modèle

le : 20 . 11 . 2017

Comparaison multiples des moyennes.

Si à la suite d'un test d'égalité de plusieurs moyennes, l'hypothèse nulle est rejetée, les moyennes présentent au moins une différence entre elles.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

Si H_0 rejetée ⇒ Au moins une moyenne diffère des autres.

– Test t.

Pour comparer les moyennes 2 à 2, les effectifs étant petits ($n < 30$), on utilise le test "t" de Student.

$$ppds = t_{(1-\alpha/2,\ \nu)} S_R \sqrt{\frac{1}{n} + \frac{1}{n'}}$$

* $t_{(1-\alpha/2,\ \nu)}$ → Table de Student.
* S_R = Variance résiduelle

$S_R = \sqrt{CM_R}$ / CM_R = carré moyen résiduelle.

Exercice : on a comparé 06 variétés d'orge d'hiver ($V_1, V_2, \ldots V_6$) en utilisant un dispositif comportant 05 blocs aléatoires complet

Bloc 1	V_1	V_6	V_4	V_3	V_5	V_2
Bloc 2	V_5	V_1	V_3	V_6	V_2	V_4
Bloc 3	V_3	V_2	V_5	V_1	V_6	V_4
Bloc 4	V_2	V_4	V_6	V_3	V_1	V_5
Bloc 5	V_1	V_2	V_4	V_5	V_3	V_6

Essai comparatif des variétés d'orge : dispositif des traitements

Essai comparatif des variétés d'orge : résultats parcellaires (Qx/ha)

							Moy
Bloc 1	63	64	64	53	60	62	61
Bloc 2	65	69	53	51	60	50	59
Bloc 3	67	54	56	52	54	50	55,5
Bloc 4	74	68	73	63	70	54	67
Bloc 5	71	70	63	62	62	68	66
Moyenne par variété	68	65	61,8	56,2	62,4	56,8	61,7

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_6$??? / on change d'origine et on prend 60 comme 0

$\sum x_i , \bar{x}_i$

$\sum x_i^2$

	V_1	V_2	V_3	V_4	V_5	V_6	$\sum x_i , \bar{x}_i$	$\sum x_i^2$	$(\sum x_i)^2/6$
Bloc 1	3	4	4	-7	0	2	6 $\bar{x}_1 = 1$	94	6
Bloc 2	5	9	-7	-9	0	-10	-6 $\bar{x}_2 = -1$	372	6
Bloc 3	7	-6	-4	-8	-6	-10	-27 $\bar{x}_3 = -4,5$	301	121,5
Bloc 4	14	8	13	3	10	-6	42 $\bar{x}_4 = 7$	574	294
Bloc 5	11	10	3	2	2	8	36 $\bar{x}_5 = 6$	302	216
$\sum x_j$	40	25	9	-19	12	-16	51		643,5
moy $\left(\bar{x} = \frac{\sum x_j}{5}\right)$	$\bar{x}_1 = 8$	$\bar{x}_2 = 5$	$\bar{x}_3 = 1,8$	$\bar{x}_4 = -3,8$	$\bar{x}_5 = 2,4$	$\bar{x}_6 = -3,2$			
$\sum x_j^2$	400	297	259	207	176	304		$\sum x_j^2 = 1643$	
$(\sum x_j)^2/5$	320	125	16,2	72,2	28,8	51,2			86,7

"Essai comparatif des variétés d'orge : Calcul des variations (origine 60)"

$$V_f = \frac{1}{k-1} \underbrace{\sum n_i (\bar{x}_i - \bar{x})^2}_{SCE} = \frac{n_A(\bar{x}_A)^2 + n_B(\bar{x}_B)^2 + n_C(\bar{x}_C)^2}{k-1}$$

$$V_f = \frac{SCE}{k-1} = C.M_T \quad (\text{carré moyen de traitement})$$

$$SCE = \sum n_i (\bar{x}_i - \bar{x})^2 = n_A(\bar{x}_A)^2 + n_B(\bar{x}_B)^2 + n_C(\bar{x}_C)^2 - N(\bar{x}).$$

$$SCE_T = 5(\bar{x}_1^2 + \bar{x}_2^2 + \bar{x}_3^2 + \cdots + \bar{x}_6^2) - N(\bar{x}) = 5(8^2 + 5^2 + 1,8^2 + 3,8^2 + 2,4^2 + 3,2^2) - 30(1,7)^2$$

$$= 5(64 + 25 + 3,24 + 14,44 + 5,76 + 10,24) - 30(1,7).$$

$$= \boxed{526,7}, \quad CM_T = \frac{SCE}{K-1} = \frac{526,7}{6-1} = \boxed{105,34}$$

pour bloc : $SCE_B = n(\bar{x}_1^2 + \cdots + \bar{x}_5^2) - N(\bar{x})^2$

$$= 6(1 + 1 + 20,25 + 49 + 36) - 30(1,7)^2 = \boxed{556,8}$$

le : 22.11.2017 :

le : 27.11.2017 :

$SCE_T = \boxed{526,7}$.

$CM_t = \frac{SCE_T}{k-1} = \frac{526,7}{5} = \boxed{105,34}$.

$SCE_{Bloc} = n(\bar{x}_1^2 + \bar{x}_2^2 + \dots + \bar{x}_5^2) - N(\bar{x})^2 = 6(1+1+20,25+49+36) - 30(1,7)^2$

$= \boxed{556,8}$.

$CM_{Bloc} = \frac{SCE_{Bloc}}{k-1} = \frac{556,8}{4} = \boxed{139,2}$

$SCE_R = \sum$ carrés résiduels

$SCE_R = (\sum x_{ij}^2 - N\bar{x}^2) - [\sum \frac{x_i^2}{5} - N\bar{x}^2 + \sum x_j^2 - N\bar{x}^2]$

$= (1643 - 86,7) - [(613 - 86,7) + (643 - 86,7)] = \boxed{473,7}$.

Analyse de la variance

$F = \frac{CM}{CM_R}$

origine de la variance	ddl	SCE	Carrés moyens CM	F	F théorique $\alpha = 5\%$
Variétés	$k-1 = 5$	526,7	105,34	4,44	2,71
Blocs	$n-1 = 4$	556,8	139,2	5,87	2,87
résiduelle	$(k-1)(n-1) = 20$	473,7	23,68		

	4	5
1		
20	2,87	2,71

Bloc : $F_{Bloc} > F_{théorique}$ → pas de différence significative
⇒ Homogénéité.

Traitement ($F_T > F_{th}$ → H_0 : rejetée ⇒ ≠ significative).

⇒ Calcul de la ppds :

$ppds = t_{[1-\frac{\alpha}{2}, (n-1)(k-1)]} \times \sqrt{CM_R\left(\frac{1}{n} + \frac{1}{n'}\right)}$

$t(0,975, 20) = 2,09$

$ppds = 2,09\sqrt{23,68\left(\frac{1}{5} + \frac{1}{5}\right)} = \boxed{6,42}$

N.B : * > ppds

	V_4 56,2	V_6 56,8	V_3 61,8	V_5 62,4	V_2 65	V_1 68
56,2	0	0,6	5,6	6,2	8,8 *	11,8 *
56,8		0	5	5,6	8,2 *	11,2 *
61,8			0	0,6	3,2	6,2
62,4				0	2,6	5,6
65					0	3

μ_4	μ_6	μ_3	μ_5	μ_2	μ_1
56,2	56,8	61,8	62,4	65	68
		61,8	62,4	65	68
			62,4	65	68
				65	68

ACP (Analyse en composante Principale).

AFC (Analyse Factorielle des composante).

le : 29.11.2017

Chapitre : Analyse en Composantes principales (ACP)

Données : n individus.

p variables.

L'ACP permet d'explorer les liaisons entre les variables et les ressemblances entre individus

- Principe de l'ACP :

on cherche une représentation des n individus, dans un sous espace F_k de R^p de dimensions k par exemple.

Autrement [...] cherche à définir k nouvelles variables qui fo[...] d'informations.

[...]ules

[...] → axe F_1.

[...] → axe F_2.

[...]tions :

[...] une projection sur un plan

[...]ales de la configuration exacte.

μ_4	μ_6	μ_3	μ_5	μ_2	μ_1
56,2	58,8	61,8	62,4	65	68
		61,8	62,4	65	68
			62,4	65	68
				65	68

ACP (Analyse en Composante Principale).

AFC (Analyse Factorielle des composante).

le : 29.11.2017

Chapitre : Analyse en Composantes principales (ACP)

Données : n individus.

p variables.

L'ACP permet d'explorer les liaisons entre les variables et les ressemblances entre individus

– Principe de l'ACP :

on cherche une représentation des n individus, dans un sous espace F_k de R^p de dimensions k par exemple.

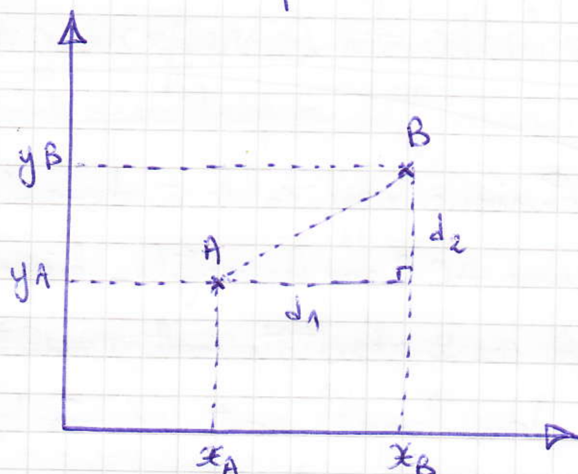
Autrement dit on cherche à définir k nouvelles variables qui feront perdre moins d'informations.

[Composantes principales
et les axes principaux → axe F_1.
→ axe F_2.

– Perdre moins d'informations :

En ACP, on cherche à obtenir une projection sur un plan présentant des distorsions minimales de la configuration exacte.

le plan de projection retenu est tel que les distances moyennes entre les points dans l'espace soient le mieux conservées.



~~$\delta^2 / / X$~~

$d^2(A,B) = d_1^2 + d_2^2$

$d^2(A,B) = (x_B - x_A)^2 + (y_B - y_A)^2$

dans l'espace R^p

$e_i(x_i^1, x_i^2, \dots x_i^p)$

$e_j(x_j^1, x_j^2, \dots x_j^p)$

$d^2(e_i ; e_j) = (x_i^1 - x_j^1)^2 + (x_i^2 - x_j^2)^2 + \dots + (x_i^p - x_j^p)^2$

$d^2(e_i ; e_j) = \sum_{k=1}^{p} (x_i^k - x_j^k)^2$

- Inertie totale :

$I_g = \sum \frac{1}{n} d^2(e_i ; g)$

ou de façon générale : $I_g = \sum p_i \, d^2(e_i ; g)$.

N.B : $\begin{pmatrix} \sigma^2 = \sum \frac{1}{n}(x_i - \bar{x})^2 \\ \sigma^2 = \sum \frac{n_i (x_i - \bar{x})}{n} \end{pmatrix}$

$I_g = p_1$.

$S_1 = \frac{1}{n}[\sum (x_1^1 - x_g)^2 + (x_2^1 - x_g)^2 + \dots + (x_p^1 - x_g)^2$

$S_2 = \frac{1}{n}[\sum (x_1^2 - x_g)^2 + (x_2^2 - x_g)^2 + \dots + (x_p^2 - x_g)^2$

$\vdots$

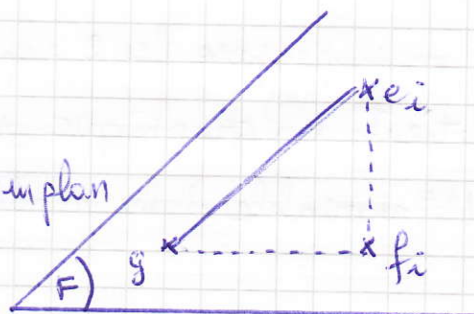
$I_g = \sum S_i^2$

L'inertie = Σ variances.

→

f_i : projection orthogonal de e_i sur un plan

g : centre de gravité



$\|e_i - g\|^2 = \|e_i - f_i\|^2 + \|f_i - g\|^2$

Le : 11.12.2017 =

Analyse factorielle des correspondances (AFC)

Introduction : c'est une technique applicable à des tableaux de contingence. Elle consiste à réaliser une ACP sur les profils-lignes et les profils-colonnes. Les résultats graphiques de ces 2 résultats sont ensuite superposés pour produire un graphique dans lequel sont réunis les modalités des 2 variables.

1- Données =

	y_1	...	y_k	....	y_c	total
x_1	n_{11}	....	n_{1k}		n_{1c}	n_{1+}
⋮						
x_l	n_{l1}		n_{lk}		n_{lc}	n_{l+}
⋮						
x_r	n_{r1}		n_{rk}		n_{rc}	n_{r+}
total	n_{+1}	....	n_{+k}	~~[illegible]~~ ...	n_{+c}	n

2- Problème =

on suppose qu'il y a une liaison entre x et y, et on cherche à décrire cette liaison.

Pour cela, on se base sur l'étude des profils-lignes et des profils-colonnes.

ex : définition du $l^{ième}$ profils-lignes.

$$\frac{n_{l1}}{n_{l+}}, \dots, \frac{n_{lk}}{n_{l+}}, \dots, \frac{n_{lc}}{n_{l+}}$$

ex : $k^{ième}$ Colonnes : $\frac{n_{1k}}{n_{+k}}, \dots, \frac{n_{lk}}{n_{+k}}, \dots, \frac{n_{rk}}{n_{+k}}$.

3. Méthode =

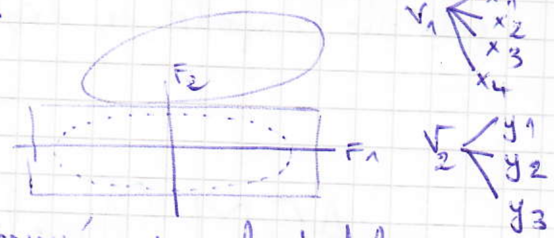
- On [illegible]lise une ACP du tableau profils-lignes (Individus),
- [illegible] " " " " profils-colonnes.
- [illegible]rpose les 2 représentations sur le même graphique
[illegible]aphique représentant à la fois les modalités
[illegible] modalités de la 2ème variable.

4- Etapes d'1 AFC :

→ nuages de points dans un espace à "p" dimensions ($p > 3$).

→ Consiste à déterminer un sous espace sur lequel on pourra projeter le nuage de points sans trop le déformer.

→ choix des axes.



5- Interprétation :

A l'aide des représentations graphiques appuyées par les tableaux, on essayera de :

- Repérer des groupes d'individus homogènes vis à vis de l'ensemble des caractères.

- Révéler des différences entre individus ou groupes d'individus, relativement à l'ensemble des caractères.

Exemple : Répartition des exploitations agricoles selon les départements et la SAU.

	INF 05	S 05 10	S 10 20	S 20 35	S 35 50	SUP 50
R1	870	870	730	680	470	890
R2	820	1260	2460	3330	2170	2960
R3	2290	1070	1420	1830	1260	2330
R4	1650	890	1350	2540	2090	3230
R5	1940	1130	1750	1660	770	1140
R6	2110	1170	1640	1500	550	430
R7	1770	820	1260	2010	1680	2090
R8	1740	920	1560	2210	1990	1240

le : 18.12.2017

F_2 (18,58)

F_1 (72,22)

total : 90,80 %

INF5

plus la surface est petite, plus elle a un score élevé sur l'axe F_1.
on peut dire que plus une région est située à droite, plus elle comporte de petites exploitations, et inversement plus elle est située à gauche plus elle comporte des exploitations de grande surface.

ACM (Analyse des correspondances multiples)

Introduction = l'AFC permet d'analyser la liaison entre 2 variables qualitatives.
Dans la pratique, en particulier dans les domaine des traitements des enquêtes (ou de questionnaires) on observe de nombreux caractères sur un nombre important d'individus.
L'ACM est la méthode factorielle qui permet de traiter ce type de problème.
L'ACM est réalisée sur un tableau de Burt relatif à au moins 3 variables qualitatives.

Application =
Soient 1635 étudiants ($n = 1635$) et 5 variables ($p = 5$).

- Sexe : Fille, Garçon.
- Serie du bac : bac A, bac B, bac C ou D, bac G, aut bac.
- L'age d'obtention du bac : 18 ans, 19 ans, 20 ans.
- Catégories socio-professionnelle des parents : art + com (artisant et commerçant); empl (employé); inter (profession intermédiaire); ouv (ouvrier); pro. lib (profession libérale); aut csp (autre catégories socio-professionnelle).
- la réussite au (moins au) DEWG : Oui - Non.